

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-104782

(43)Date of publication of application : 21.04.1995

(51)Int.Cl. G10L 3/00

(21)Application number : 05-247836

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN
KENKYUSHO:KK

(22)Date of filing : 04.10.1993

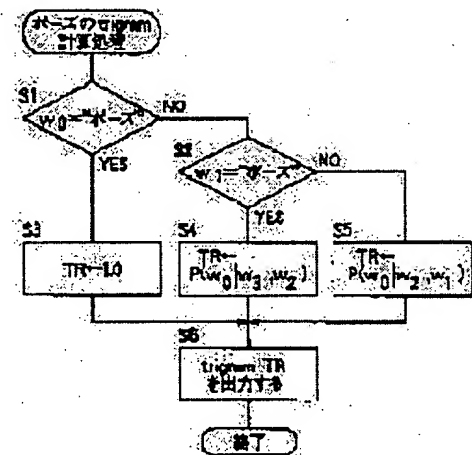
(72)Inventor : MURAKAMI JINICHI

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PURPOSE: To provide a voice recognition device which can obtain a high sentence recognition rate as compared with a conventional example even when an inputted spoken voice includes a pause or redundant word.

CONSTITUTION: The voice recognition device which performs voice recognition by calculating a probability value of a word to be recognized in the spoken voice sentence consisting of a character string inputted by referring to a specific statistical language model on the basis of one or plural words connected in front of the word sets the probability value of the language model of the pause in the spoken voice sentence or the word connected to the redundant word to 1 or a value close to 1, and calculates the probability of the language model of the work by skipping the pause or redundant word connected to a work other than the pause or redundant work across the pause or redundant work, thereby performing the voice recognizing process.



LEGAL STATUS

[Date of request for examination]

08.05.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(51) Int.Cl.⁶

G 1 0 L 3/00

識別記号

5 3 1 P 9379-5H

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 5 O L (全 7 頁)

(21) 出願番号 特願平5-247836
 (22) 出願日 平成5年(1993)10月4日

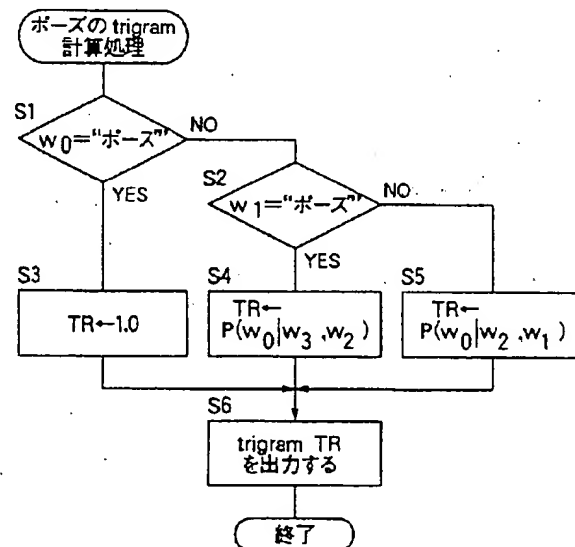
(71) 出願人 593118597
 株式会社エィ・ティ・アール音声翻訳通信
 研究所
 京都府相楽郡精華町大字乾谷小字三平谷5
 番地
 (72) 発明者 村上 仁一
 京都府相楽郡精華町大字乾谷小字三平谷5
 番地 株式会社エィ・ティ・アール音声翻
 訳通信研究所内
 (74) 代理人 弁理士 青山 葆 (外2名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【目的】 入力された発声音声にポーズや冗長語があった場合であっても、従来例に比較して高い文認識率を得ることができる音声認識装置を提供する。

【構成】 所定の統計的言語モデルを参照して入力された文字列からなる発声音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、上記発声音声文内のポーズ又は冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズ又は冗長語を介してポーズ及び冗長語以外の単語に接続されるとき上記ポーズ又は冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する。



【特許請求の範囲】

【請求項1】 所定の統計的言語モデルを参照して入力された文字列からなる発声音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、

上記発声音声文内のポーズ又は冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズ又は冗長語を介してポーズ及び冗長語以外の単語に接続されるとき上記ポーズ又は冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する音声認識手段を備えたことを特徴とする音声認識装置。

【請求項2】 所定の統計的言語モデルを参照して入力された文字列からなる発声音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、

上記発声音声文内のポーズに接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズを介してポーズ以外の単語に接続されるとき上記ポーズをスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する音声認識手段を備えたことを特徴とする音声認識装置。

【請求項3】 所定の統計的言語モデルを参照して入力された文字列からなる発声音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、

上記発声音声文内の冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、冗長語を介して冗長語以外の単語に接続されるとき上記冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する音声認識手段を備えたことを特徴とする音声認識装置。

【請求項4】 上記所定値は、0.8以上1.0以下であることを特徴とする請求項1、2又は3記載の音声認識装置。

【請求項5】 上記統計的言語モデルは、単語のtriggerであることを特徴とする1、2、3又は4記載の音声認識装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、発声音声におけるポーズ及び冗長語について処理を行う音声認識装置に関する。

【0002】

【従来の技術】 近年、連続音声認識の研究が盛んに行われ、いくつかの研究機関で音声認識システムが構築されている。これらのシステムの多くは丁寧に発声された

音声を入力対象にしている。しかしながら、人間同士のコミュニケーションでは、「あのー」、「えーと」などに代表される冗長語や、一時的に発聲音声が無い状態（以下、ポーズという。）である言い淀みや言い誤り及び言い直しなどが頻繁に出現する。

【0003】

【発明が解決しようとする課題】 一方、連続音声認識に利用できるアルゴリズムの1つにOne pass DP (Viterbi search) アルゴリズムがある。このアルゴリズムを用いた音声認識装置においては、入力された発聲音声にポーズや冗長語があった場合、文認識率が低下するという問題点があった。

【0004】 本発明の目的は以上の問題点を解決し、入力された発聲音声にポーズや冗長語があった場合であっても、従来例に比較して高い文認識率を得ることができ音声認識装置を提供することにある。

【0005】

【課題を解決するための手段】 本発明に係る請求項1記載の音声認識装置は、所定の統計的言語モデルを参照して入力された文字列からなる発聲音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、上記発聲音声文内のポーズ又は冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズ又は冗長語を介してポーズ及び冗長語以外の単語に接続されるとき上記ポーズ又は冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する音声認識手段を備えたことを特徴とする。

【0006】 また、請求項2記載の音声認識装置は、所定の統計的言語モデルを参照して入力された文字列からなる発聲音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、上記発聲音声文内のポーズに接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズを介してポーズ以外の単語に接続されるとき上記ポーズをスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する音声認識手段を備えたことを特徴とする。

【0007】 さらに、請求項3記載の音声認識装置は、所定の統計的言語モデルを参照して入力された文字列からなる発聲音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、上記発聲音声文内の冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、冗長語を介して冗長語以外の単語に接続されるとき上記冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する音声認識手段を備

えたことを特徴とする。

【0008】また、請求項4記載の音声認識装置は、請求項1、2又は3記載の音声認識装置において、上記所定値は、0.8以上1.0以下であることを特徴とする。さらに、請求項5記載の音声認識装置は、請求項1、2、3又は4記載の音声認識装置において、上記統計的言語モデルは、単語のtrigramであることを特徴とする。

【0009】

【作用】請求項1記載の音声認識装置においては、上記音声認識手段は、上記発声音声文内のポーズ又は冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズ又は冗長語を介してポーズ及び冗長語以外の単語に接続されるとき上記ポーズ又は冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する。

【0010】また、請求項2記載の音声認識装置においては、上記音声認識手段は、上記発声音声文内のポーズに接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズを介してポーズ以外の単語に接続されるとき上記ポーズをスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する。

【0011】さらに、請求項3記載の音声認識装置においては、上記音声認識手段は、上記発声音声文内の冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、冗長語を介して冗長語以外の単語に接続されるとき上記冗長語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行する。

【0012】

【実施例】以下、図面を参照して本発明に係る実施例の音声認識装置について説明する。図1の本実施例の音声認識装置において、音素照合部4は、入力される発声音声に関するデータに基づいて、隠れマルコフモデル（以下、HMMという。）メモリ5内の音響モデルであるHMMを参照して冗長語やポーズを認識した後、One pass DPアルゴリズムを用いたOne pass DP音声認識部（以下、音声認識部という。）6は、統計的言語モデルメモリ7内の統計的言語モデル（単語のtrigram）を参照して音声認識を実行するときに、冗長語又はポーズをスキップすることにより、冗長語及び／又はポーズを含んだ音声进行を認識することを特徴とする。

【0013】まず、文音声認識アルゴリズムの改良について述べる。上記のOne pass DPアルゴリズムの経路計算において、最尤の単語列を得るために2つの方法がある。

(1) トレースバック：各時刻・各状態において、最大累積尤度を計算したときに、選択した経路を記憶してお

く。そして尤度計算の終了後、上記記憶された経路に沿ってトレースバックを行なうことによって音声認識処理を実行する（以下、第1の経路計算方法という。；例えば、Kai-Fu Lee, "Large-Vocabulary Speaker Independent Continuous Speech Recognition: The SPHINX System", 15213 CMU-CS-88-148, 1988年4月18日参照。）。)

(2) 最大累積尤度と同時に計算する：各時刻、各状態において、最大累積尤度を計算したときに、同時に、選択した経路を次の状態に渡すことによって音声認識処理を実行する（以下、第2の経路計算方法という。；例えば、村上仁一, "単語のtrigramを用いた連続音声認識の1アルゴリズム", 音響学会講演論文集, p. 185-186, 2-Q-7, 1992年10月参照。）。)

【0014】上記第1の経路計算方法は、計算量が少なく済むため従来からしばしば利用されている。一方、上記第2の経路計算方法は、第1の経路計算方法と比較すると計算量が増加するが、各時刻・各状態においてトレースバックをしなくても経路を知ることが可能であるため、言語モデルにおけるleft-right型のLRパーザと組み合わせることが容易であり、また、多くの場合、第1の経路計算方法と比較して少量のメモリで済むので、本実施例の音声認識部6は後者の第2の経路計算方法を採用する。

【0015】以下、本実施例の音声認識方法を用いる音声認識装置を示す図1を参照して、本実施例の統計的言語モデルを用いた音声認識装置の構成及び動作について説明する。

【0016】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、 Δ 対数パワー及び16次 Δ ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して音素照合部4に入力される。音素照合部4に接続される隠れマルコフモデル（以下、HMMという。）メモリ5内のHMMは、複数の状態と、各状態間の遷移を示す弧から構成され、各弧には状態間の遷移確率と入力コードに対する出力確率を有している。音素照合部4は、入力されたデータに基づいて音素照合処理を実行して音素データを、音声認識部6に出力する。

【0017】単語のtrigramを含む所定の統計的言語モデルを予め記憶する統計的言語モデルメモリ7は音声認識部6に接続される。音声認識部6は、統計的言語モデルメモリ7内の統計的言語モデルを参照して、所定のOne pass DPアルゴリズムを用いて、入力された音素データについて左から右方向に、後戻りな

5

しに処理してより高い生起確率の単語を音声認識結果データと決定することにより音声認識の処理を実行して、決定された音声認識結果データ（文字列データ）を出力する。

【0018】次いで、音声認識部6におけるポーズの処理について説明する。ポーズは文節間において出現することが多いが、音声のあらゆる場所に出現する可能性がある。しかしながら、従来の言語モデルでは、これに追従できないため、ポーズの区間で誤認識が起きやすい。そこで、単語のtrigramなる統計的言語モデルと、One pass DPアルゴリズムを用いて、全ての単語と単語の境界にポーズが入力されても文認識が可能な音声認識方法を発明した。本発明に係る実施例においては、まずポーズを1単語と考えて、ポーズに接続される単語のtrigramの値は1.0にする。そしてポーズ以外の単語に接続されるときポーズをスキップして単語のtrigramを計算する。例えば「東京都港区 新橋/pause (ポーズ) /1丁目」という文字列なる文が入力されたとき、その確率を、 $P(\text{新橋} | \text{東京都 港区}) \times 1.0 \times P(1\text{丁目} | \text{港区 新橋})$ と計算する。ここで、 $P(A | B)$ は、単語Bの後に単語Aが来る確率であり、以下、同様である。このようにすると、近似解ではあるが、ポーズを除いて単語trigramを用いたときの最尤の解が得られる。

【0019】以上の実施例において、ポーズに接続される単語のtrigramの値を1.0としているが、本発明はこれに限らず、ポーズに接続される単語のtrigramの値を好ましくは0.8以上1.0以下の範囲のある値に設定する。

【0020】図2は、図1の音声認識装置において実行されるポーズのtrigram計算処理を示すフローチャートである。なお、冗長語についても図2のフローと同様に処理される。当該計算処理は、単語列 w_3, w_2, w_1, w_0 が入力されたときに単語 w_0 のtrigramを計算する方法であって、図2に示すように、まず、ステップS1において単語 w_0 がポーズであるか否かが判断され、ステップS2において、タンゴ w_1 がポーズであるか否かが判断される。ステップS1においてYESであれば、ステップS3において単語 w_0 のtrigramの値TRを1.0に設定した後、ステップS6に進む。ステップS1においてNOの場合は、ステップS2に進み、ステップS2においてYESであれば、ステップS4において確率値 $P(w_0 | w_3, w_2)$ を単語 w_0 のtrigramの値TRとして設定した後、ステップS6に進む。一方、ステップS2においてNOであれば、ステップS5において確率値 $P(w_0 | w_2, w_1)$ を単語 w_0 のtrigramの値TRとして設定した後、ステップS6に進む。ステップS6においては、設定された単語 w_0 のtrigramの値TRを計算値として出力して当該計算処理を終了する。

6

【0021】さらに、音声認識部6における冗長語の処理について説明する。自由発話では「あの一」、「えーと」などの冗長語が多く出現する。本発明者が、詳細後述するテストデータから収集した自由発話における出現回数が2以上の冗長語を表1乃至表3に示す。

【0022】

【表1】

冗長語	出現回数
「あ」	604
「あー」	268
「あーっと」	2
「あーん」	5
「ああ」	7
「あつ」	151
「あの」	1809
「あの一」	2025
「あのを」	77
「あの一ー」	3
「い」	26
「いー」	58
「いやー」	2
「う」	23
「うー」	71
「うーん」	26
「うーんと」	2
「うん」	7
「え」	1040
「えー」	3105
「えーっと」	256
「えーっとー」	2

【0023】

【表2】

冗長語	出現回数
「えーっとですね」	8
「えーと」	466
「えーとー」	4
「えーとですね」	3
「えーまあ」	3
「えーん」	2
「ええ」	13
「えつ」	22
「えっーと」	4
「えつと」	62
「えつとー」	11
「えと」	47

	7
「えとー」	13
「お」	59
「おー」	196
「おっ」	2
「こう」	9
「この」	9
「このー」	4
「じゃ」	4
「す」	8
「すー」	2

【0024】

【表3】

冗長語	出現回数
「すっ」	2
「そ」	2
「その」	115
「そのー」	48
「ちょっと」	8
「つ」	2
「で」	61
「でー」	13
「と」	77
「とー」	11
「は」	4
「はあー」	2
「ふーん」	2
「ま」	263
「まー」	8
「まあ」	186
「まあ」	176
「まっ」	5
「も」	2
「ん」	27
「んー」	19
「んと」	2

【0025】この冗長語は文の全ての場所に出現する可
能性があるという点でポーズと似た性質がある。従っ
て、冗長語の処理にはポーズの処理と同様な手法が使用
できる。すなわち、音響モデルでは冗長語を認識しなが
ら、言語モデルでは冗長語をスキップする。例えば「東
京都 港区 新橋 えーと 1丁目」という文字列なる
文が入力されたとき、その確率値を、 $P(\text{新橋} | \text{東京
都 港区}) \times 1.0 \times P(1\text{丁目} | \text{港区 新橋})$ と計算
する。従って、冗長語の処理においても、図2に示した
フローチャートにおける「ポーズ」を「冗長語」と置き
換えて同様に処理することができる。

【0026】以上の実施例において、冗長語に接続され
る単語のtrigramの値を1.0としているが、本
発明はこれに限らず、冗長語に接続される単語のtri
gramの値を好ましくは0.8以上1.0以下の範囲
のある値に設定する。

【0027】本発明者は、以上説明した本実施例の音声
認識装置を用いて文認識率による評価を行うためにシミュ
レーションを行った。言語情報には、単語のbigram
又はtrigramを用いて特定話者及び不特定話
者の音声認識シミュレーションを行った。テストデータ
はナレータが発声した国際会議の問い合わせの文（いわ
ゆる、モデル会話。）を使用した。なお、テストデータ
の先頭及び末尾には約20ミリ秒のポーズがある。ま
た、trigramの連鎖確率値は、本出願人の対話デ
ータベースの中から国際会議の予約に関するデータであ
る約12,000文章（約170,000単語）に、テ
ストデータのテキストを加えて計算した。

【0028】上述のポーズの処理を行わないシミュレー
ションにおいては、trigramの場合、特定話者で
は78.6%の文認識率が得られる一方、不特定話者で
は59.5%の文認識率が得られた。これに対して、図
2に示したポーズの処理を実行した文認識率のシミュ
レーションにおいては、特定話者認識において86.3%
のより高い文認識率が得られ、音声認識性能が向上し
た。一方、不特定話者認識においては60.3%の文認
識率が得られ、音声認識性能の向上は顕著ではないが、
若干向上した。

【0029】さらに、自由発話の文認識シミュレーショ
ンについて述べる。ここでは、特に、冗長語の処理の有
効性を自由発話の音声で調べる。自由発話の定義は人によ
って異なるが、本実施例を用いたシミュレーションでは
以下に示すような方法で収録した音声データを使用す
る。

(1) 朗読音声データ：テキストを読みあげた音声デー
タであって、冗長語や、言い淀み、言い直しは無い。

(2) 疑似自由発話データ：冗長語をつけてテキストを
読み上げた音声データである。冗長語を除いて、上記

(1) 朗読発声データと発話内容は同一であって、言い
淀み、言い直しは無い。

(3) 自由発話データ：話者はテキストを覚えて、その
意図を理解し、自由に発話した音声データである。発話
内容は上記(1)朗読発声データと異なる。ここで、言
い淀み、言い直しは無い。

【0030】話者は、ナレータではなく、一般の人であ
る。冗長語として、「あー」「えーと」「まあ」など
の109種類を定義する。音声認識シミュレーション
は、不特定話者認識の単語のtrigramの場合のみ
行った。また、上述のポーズの処理も行った。このシミュ
レーション結果を表4に示す。表4から明らかなよう
に、冗長語を付けた疑似自由発話の音声では、64.4

%述べる認識率が得られる一方、自由発話音声では3 *【0031】

4. 4%の認識率が得られた。 *【表4】

自由発話の文認識シミュレーション結果——文認識率(%)

発話様式	+ポーズ処理	+ポーズ処理+冗長語処理
朗読発声	82.6%	74.8%
疑似自由発話	26.7%	64.4%
自由発話	14.1%	34.4%

【0032】従って、上述の冗長語の処理は、疑似自由発話でも自由発話でも、ポーズの処理のみの方法より有効であること、並びに、朗読発声の文認識においても従来の方法と比較して認識率があまり低下しないこと(82.6%→74.8%)を考慮すると、自由発話の認識において有効であることがわかる。

【0033】以上の実施例において、音声認識部6における冗長語の処理及びポーズの処理について説明しているが、音声認識部6は、冗長語の処理とポーズの処理とのうち少なくとも一方を含むように構成してもよい。

【0034】

【発明の効果】以上詳述したように本発明によれば、所定の統計的言語モデルを参照して入力された文字列からなる発声音声を、音声認識すべき単語の確率値をその単語の前に接続される1個又は複数個の単語に基づいて計算することによって音声認識する音声認識装置において、上記発声音声文内のポーズ又は冗長語に接続される単語の言語モデルの確率値を1又は1に近い値である所定値とする一方、ポーズ又は冗長語を介してポーズ及び冗長語以外の単語に接続されるとき上記ポーズ又は冗長

語をスキップして単語の言語モデルの確率値を計算して音声認識処理を実行するように構成したので、ポーズ又は冗長語を考慮して音声認識することができ、これによって、認識率が大幅に向上し、音声認識性能を向上させることができる。

【図面の簡単な説明】

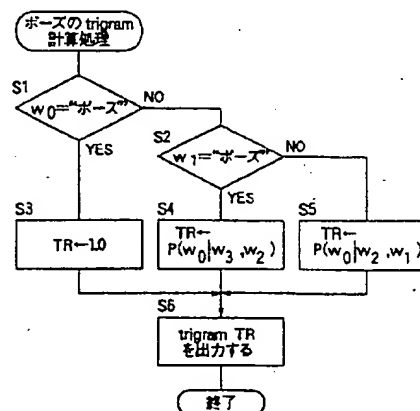
【図1】 本発明に係る一実施例である音声認識装置のブロック図である。

【図2】 図1の音声認識装置において実行されるポーズのtrigram計算処理を示すフローチャートである。

【符号の説明】

- 1…マイクロホン、
- 2…特徴抽出部、
- 3…バッファメモリ、
- 4…音素照合部、
- 5…One pass DP音声認識部、
- 6…隠れマルコフモデル(HMM)メモリ、
- 7…統計的言語モデルメモリ。

【図2】



【図1】

